

Weekly Report

2013.5.27-2013.6.2

黄芯芯

本周工作：

1. 配置浙一项目的工程环境，调试读取 DCM 数据程序的代码，查找头颅数据的断层问题。
2. 阅读论文《Narrative Visualization: Telling Stories with Data》，受文章启发觉得可以做一个关于 NBA 比赛数据可视化的项目，框架大概是每天爬去 NBA 比赛的各种数据，对这些数据按照一场比赛的时间做时序的可视化，让用户去发现和挖掘决定比赛胜负的因素等。这个可视化的侧重点在于讲述一场比赛的故事，这也是和现有的一般的 NBA 新闻网站上只有统计数据的区别之处。另外，在周三的组会讨论后，又汇集了更多的想法。
3. 审基金。
4. 关于淘宝第一大食品店数据的讨论：
 - 1) 首先，大家都觉得这个数据可以挖掘的研究的信息是很多的，大家也 yy 了各种各样的想法：对卖家而言，他们应该想知道在什么时间段销售什么商品，做什么促销活动会更有利于提高销量；包邮对销量的影响；买家的时序购买特征、地域购买特征；影响交易成功或关闭的因素；某些维度对另外一些维度的影响等等。
 - 2) 虽然数据维度很多，但是诸如名称、手机号码、账号等一些维度是没有意义的。真正使用数据的时候需要大规模的清洗和加密。
 - 3) 数据量应该可以很大，所以时序信息不是问题。

下周计划：

1. 肝脏项目答辩。
2. 写一下 NBA 数据可视化的框架和规划。
3. 继续调试浙一的项目代码。